Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 7144–7148, August 1992
Biochemistry

At least five related, but distinct, hepatitis C viral genotypes exist

(non-A, non-B hepatitis/reverse-transcription-PCR/direct double-stranded sequencing/genotype-specific probes)

T.-A. CHA*, E. BEALL, B. IRVINE, J. KOLBERG, D. CHIEN, G. KUO, AND M. S. URDEA

Chiron Corporation, 4560 Horton Street, Emeryville, CA 94608

Communicated by William J. Rutter, April 29, 1992 (received for review February 4, 1992)

ABSTRACT **Hepatitis C virus, the major causative agent of blood-borne non-A, non-B hepatitis in the world, has been the subject of considerable nucleic acid sequence analysis. Although all reported hepatitis C sequences from the United States have been represented by the prototype hepatitis C virus type 1 sequence, two groups of variant sequences have been reported in Japan. However, we have noted five distinct, but related, genotypes (I–V) throughout the world, based on detailed sequence determination and analysis of the first 1700 nucleotides and part of the nonstructural region 5 at the C terminus of the open reading frame. The nucleotide sequence for a large number of hepatitis C virus isolates spanning six continents was obtained by direct sequence analysis of PCR products after reverse transcription. Genotype was classified by using several distinct sequence motifs. We observed that most genotypes coexist in several geographic regions, including the United States, Japan, Germany, and Italy. So far, genotype V has been found only in South Africa. Interestingly, each distinct genotype seems to be maintained throughout the genome in the segments studied. These genotype distinctions should be considered when designing specific diagnostic tests, developing potential vaccines, and studying viral transmission.**

Hepatitis C virus (HCV) was isolated from a chimpanzee chronically infected with a contaminated human factor VIII concentrate in 1989 (1). An immunoassay for circulating antibody to HCV was developed by using a yeast-derived recombinant antigen encoded by a segment of the cloned HCV genome (2). Results indicated HCV to be the major causative agent of the blood-borne non-A, non-B hepatitis (1–3).

The genomic organization and characterization of HCV has been reported (4, 5). The HCV genome is a positive-strand RNA of ≈9.4 kilobases and contains a large open reading frame that encodes a polyprotein precursor of 3011 amino acids. Based on comparative sequence analysis of the genome and encoded polyprotein, HCV is thought to be distantly related to the flaviviruses and the pestiviruses. The large open reading frame appears to encode colinearly structural and nonstructural proteins, with the structural proteins located at the 5′-end portion of the genome. Putative boundaries are assigned that separate the 5′-untranslated region (5UT), the core protein (C), the glycoproteins envelope 1 (E1) and nonstructural protein 1/envelope 2 (NS1/E2), the nonstructural proteins 2–5 (NS2–NS5), and the 3′-untranslated region (3UT). The prototype HCV nucleotide sequence reported is termed HCV1 (refs. 4 and 5; see ref. 6 for review).

Using pooled plasma samples from human non-A, non-B hepatitis patients and potential HCV carriers, Kato *et al.* (7), Takamizawa *et al.* (8), and Okamoto *et al.* (9) have reported entire genomic sequences of Japanese HCV strains. Some partial 5′-end HCV sequences from individual Japanese samples were reported by Takeuchi *et al.* (10, 11) and Okamoto *et al.* (12). In addition, partial sequences in the NS3, NS4, and NS5 regions have been reported by several Japanese groups (13–16). It is significant that most Japanese HCV sequences reported varied markedly from HCV1, suggesting multiple strains or distinct genotypes of HCV.

Despite numerous HCV sequences reported from Japan and the United States, information about HCV nucleotide sequence in other parts of the world is scarce. In the literature, the terminology used to distinguish different genetic variants of HCV is not consistent, and classification based on an average percent homology is ambiguous. In this study, we term the genetic variants of HCV as genotypes and classify them by the specific sequence motifs that comprise distinctive variation patterns when compared with the HCV1 prototype sequence. We sequenced the 5′-end region (5UT, C, E1, and N terminus of NS1/E2) and a portion of the NS5 domain from geographically diverse HCV isolates and found at least five related, but distinct, HCV genotypes.†

MATERIALS AND METHODS

HCV-Containing Serum Samples. A total of 72 serum samples were collected from geographically diverse areas including the United States, Japan, Australia, Argentina, Spain, Germany, Italy, and South Africa. The American and Japanese samples (except one, NAC5) were from blood donors, and all others were from non-A, non-B hepatitis patients. All serum samples were positive in an immunoassay against various recombinant antigens encoded by the HCV genome (G.K., unpublished results). Two samples with known sequences, HCV1 (4, 5) and J1 (10, 11), were also included in our studies. Our sequence data for these two isolates were consistent with published results.

Viral RNA Isolation. RNA was extracted from serum with guanidinium salt, phenol, and chloroform (ref. 17; RNAzol B, Cinna/Biotecx Laboratories, Friendswood, TX). Extracted RNA was precipitated with isopropanol and washed with ethanol. A total of 25 μl of serum was processed for RNA isolation, and the purified RNA was resuspended in 5 μl of diethyl pyrocarbonate-treated H_2O for subsequent cDNA synthesis.

cDNA Synthesis and PCR Amplification. For cDNA synthesis and PCR amplification, we used a protocol developed by Perkin–Elmer/Cetus (GeneAmp RNA PCR kit) without alteration. Both random hexamer and primers with specific complementary sequences to HCV (Table 1) were used to prime the reverse transcription (RT) reaction. All processes, except for adding and mixing reaction components, were done in a thermal cycler (MJ Research, Watertown, MA). The first-strand cDNA synthesis reaction was inactivated at

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: HCV, hepatitis C virus; 5UT, 5′-untranslated region; C, core protein; E1 and E2, envelope proteins 1 and 2; NS1–NS5, nonstructural protein 1–5; RT, reverse transcription; GI–GV, genotype I–V.

*To whom reprint requests should be addressed.

†The sequences reported in this paper have been deposited in the GenBank data base (accession nos. M74804–M74815).

Table 1. Oligonucleotide primers and probes

Sequence name*	Sequence (5′ → 3′)†	Nucleotide position‡	Use§
14s	AACTACTGTCTTCACGCAGAAAGC	53–76	PCR primer
30s	ACTTCCGAGCGGTCGCAACCT	495–515	PCR primer
39s	GCCGACCTCATGGGGTACATA	732–752	PCR primer
40s	CAAACGTAACACCAACCGRCGCCCACAGG	374–402	PCR primer
45s	CGACGTCACATCGATCTGCTTGT	1116–1138	PCR primer
46s	CGACGCCACGTCGACTTGCTCGT	1116–1138	PCR primer
48s	ATGATGAACTGGTCVCCYAC	1308–1327	PCR primer
90s	CCATGAATCACTCCCCTGTGAGGAACTA	30–57	PCR primer
14a	GGGCCCCAGCTAGGCCGAGA	667–648	PCR primer
45a	TCATCATATCCCAWGCCA	1311–1294	PCR primer
48a	AGCTGAGCCATTACCAACGC	1351–1332	cDNA synthesis and PCR primer
49a	AACTGCGACACCACTAAGGC	1351–1332	cDNA synthesis and PCR primer
50a	ACAGAYCCGCAKAGRTCCCCCACG	1192–1169	PCR primer
51a	CTGTTGCATAGTTCACGCCGTCYTCCAGA	840–812	PCR primer
58a	TGATRGGGCCCCAKCCCTGGTCRAA	1758–1734	cDNA synthesis and PCR primer
61a	ACCTTVGCCCAGTTSCCCRCCATGGA	1453–1428	PCR primer
S-primer	TGGGGATCCCGTATGATACCCGCTGCTTTGA	8245–8275	PCR primer (16)
AS-primer	GGCGGAATTCCTGGTCATAGCCTCCGTGAA	8645–8616	PCR primer (16)
GIC	GCAACCTCGAGGTAGACGTCAGCCTATCCC	509–538	GI C probe
GIIC	GCAACCTCGTGGAAGGCGACAACCTATCCC	509–538	GII C probe
GIE	GTCACCAATGATTGCCCTAACTCGAGTATT	948–977	GI envelope probe
GIIE	GTCACGAACGACTGCTCCAACTCAAG	948–973	GII envelope probe

*s, Sense, a, antisense.
†R = A, G; V = A, C, G; Y = C, T; W = A, T; K = G, T; S = G, C.
‡Reference is prototype HCV1 (4, 5).
§All PCR primers except two (48a and 49a) were also used as sequencing primers.

99°C for 5 min and then cooled at 5°C for 5 min before adding reaction components for subsequent amplification. After an initial five cycles of 97°C for 1 min, 50°C for 2 min, and 72°C for 3 min, 30 cycles of 94°C for 1 min, 55°C for 2 min, and 72°C for 3 min followed and then a final 7 min of elongation at 72°C. In general, a nested second round of PCR amplification was done by using 1% of the first PCR product as the template to obtain sufficient specific PCR product for sequencing.

DNA Sequencing. The specific PCR products were isolated from 1.5% agarose gels and purified with Prep-A-Gene (Bio-Rad). The purified double-stranded DNA was melted at 95°C for 5 min and then quickly chilled on ice. Primers used to generate the PCR products were used as sequencing primers to sequence both strands. Both *Taq* polymerase (Perkin–Elmer/Cetus) and *Bst* polymerase (Bio-Rad) were used in the dideoxynucleotide chain-termination sequencing reactions according to the manufacturers' instructions.

Oligonucleotide Probe Hybridization. Four genotype-specific oligonucleotide probes were designed for the HCV C and E1 regions. Their sequences and positions in the HCV genome are listed in Table 1. The PCR products were first analyzed by agarose gel electrophoresis. Four identical gels containing the PCR products generated by the 40s and 50a primers (Table 1) were run. The gels were then blotted onto nylon membranes and hybridized with a ^{32}P-labeled oligonucleotide probe. We found that incorporation of 3 M tetramethylammonium chloride in the hybridization solution enhanced specificity of probe binding to the specific genotype. Hybridization was done at 50°C overnight, and specific washing conditions are described in the legend for Fig. 3. The procedures for hybridization and Southern blotting onto nylon membranes with tetramethylammonium chloride were according to Sambrook *et al.* (18).

Computer Programs for Sequence Analysis. The PCR primers used for the 5′-end portion of HCV genome were designed on the basis of published HCV1 (4, 5) and J1 (10, 11) sequences and optimized with the program OLIGO (National Biosciences, Hamel, MN). GENALIGN was used for sequence alignments and comparisons (IntelliGenetics).

RESULTS

Sequencing Strategy. RNA viruses commonly exist as heterogeneous populations of nonunique but related genomes (19). When PCR products were cloned into M13-based vectors for sequencing, sequence differences between clones for the same isolate were often noted—most likely due to population heterogeneity. Incorporation errors by *Taq* polymerase during PCR also contribute to this heterogeneity (20, 21). We chose to sequence the PCR products directly without additional cloning steps. If an individual had multiple strands of HCV, only the dominant sequence would be determined.

To verify the accuracy of sequences, several overlapping regions produced by PCR were sequenced. All primer sequences and their positions in the HCV genome are listed in Table 1. Overlapping PCR fragments generated by using these primers at the 5′-end of the HCV genome appear in Fig.

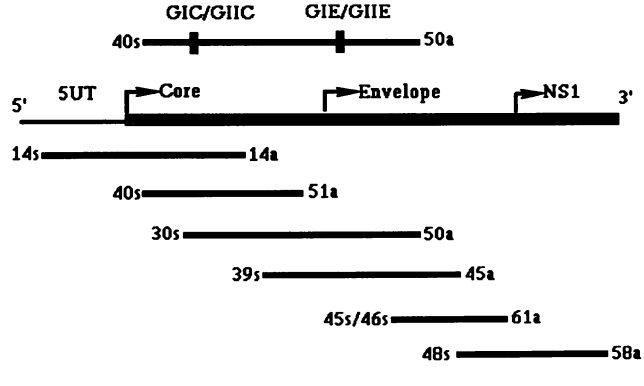

FIG. 1. Schematic representation of sequencing strategy. The first 1800-nucleotide region of the HCV genome is shown graphically. Relative positions of 5UT, C, E, and NS1 domains are marked. The various overlapping PCR fragments for sequencing are shown below. Above the HCV genome is shown the PCR fragment (40s/50a) for probing with genotype-specific oligonucleotide probes. E and NS1 domains are also referred to as E1 and E2/NS1, respectively (6).

1. Samples were meticulously handled to reduce contamination. Water and normal human serum were used as controls to monitor false positive results. In addition, the entire process of RNA extraction, RT–PCR, and sequencing was done multiple times on different occasions for each HCV isolate. Obtaining matched overlapping sequences for a single isolate was considered evidence for authenticity of the sequence determined for that isolate. Nonetheless, there was an average 0.2% combined sequence uncertainty for all isolates in our study. When a nucleotide identity was uncertain, the HCV1 sequence (4, 5) was assigned for alignment purposes. The minor uncertainty in the sequence determination does not lead to ambiguity in genotype I–V (GI–GV) assignment, which is based on several stretches of sequence motifs (see below).

HCV Sequence Exhibits Multiple Genotypes. We propose to define the genetic variants of HCV as different genotypes based on specific sequence motifs distinct from the HCV1 prototype sequence. Fig. 2 displays some representative variation patterns in the four domains (5UT, C, E1, and NS5) examined. Sequence divergence is not evenly distributed over different domains; however, distinct genotypic sequences seem to be maintained throughout the entire HCV genome. It is imperative to examine longer segments of sequence for each isolate in several domains to discern different genotypes. For instance, isolates US4 and US5 cannot be assigned to different genotypes based on the 5UT region alone (Fig. 2*A*), unless other domains (Fig. 2 *B* and *C*) are taken into consideration. Isolates US22 and US23 are assigned to a single genotype (GIV) based on the 5UT region (Fig. 2*A*), which is consistent with sequence motifs found in the C (Fig. 2*B*) and NS5 (Fig. 2*D*) domains. Furthermore, isolates SA3 and SA4 are tentatively assigned to a genotype (GV) based on the NS5 domain alone (Fig. 2*D*), even though their sequences in the 5UT region are almost identical to that of GI and GII isolates (Fig. 2*A*). We note that the 5UT region has an overall 94–100% nucleotide homology among all genotypes. With the above definition, the sequences of most published United States (4, 5, 22) and Japanese isolates (7, 8, 10–12) are classified as GI and GII, respectively. GIII sequences are also reported for some Japanese isolates (9, 14, 16).

It has been suggested that primers designed in the 5UT region are extremely specific, and their use for detecting HCV RNA in serum samples has been demonstrated (23). For the 72 serum samples examined here, 57 were found positive by the RT–PCR method with these 5UT primers. However, only 44 samples were positive with another primer pair (48s/61a, Table 1) located in the E1 region. By probing this fragment with GI- and GII-specific probes, a total of 35 samples were identified as either genotype. We sequenced this fragment from four of the remaining nine isolates and found that they clearly exhibited different nucleotide patterns from GI and GII (Fig. 2*C*). We have designated two additional genotypes based on this data, and we have assigned one isolate to GIII (isolate I10) and the other three isolates to GIV (isolate GH8, I4, and I11).

Enomoto *et al.* (16) have reported efficient use of PCR primers designed in the NS5 region to detect HCV in serum samples. By sequencing many isolates in this region, they have identified two "types" of HCV sequences from Japanese isolates, termed K1 and K2. By using their primers, we obtained partial NS5 sequence for a total of 22 isolates. Based on sequence motifs, a total of five genotypes can be identified

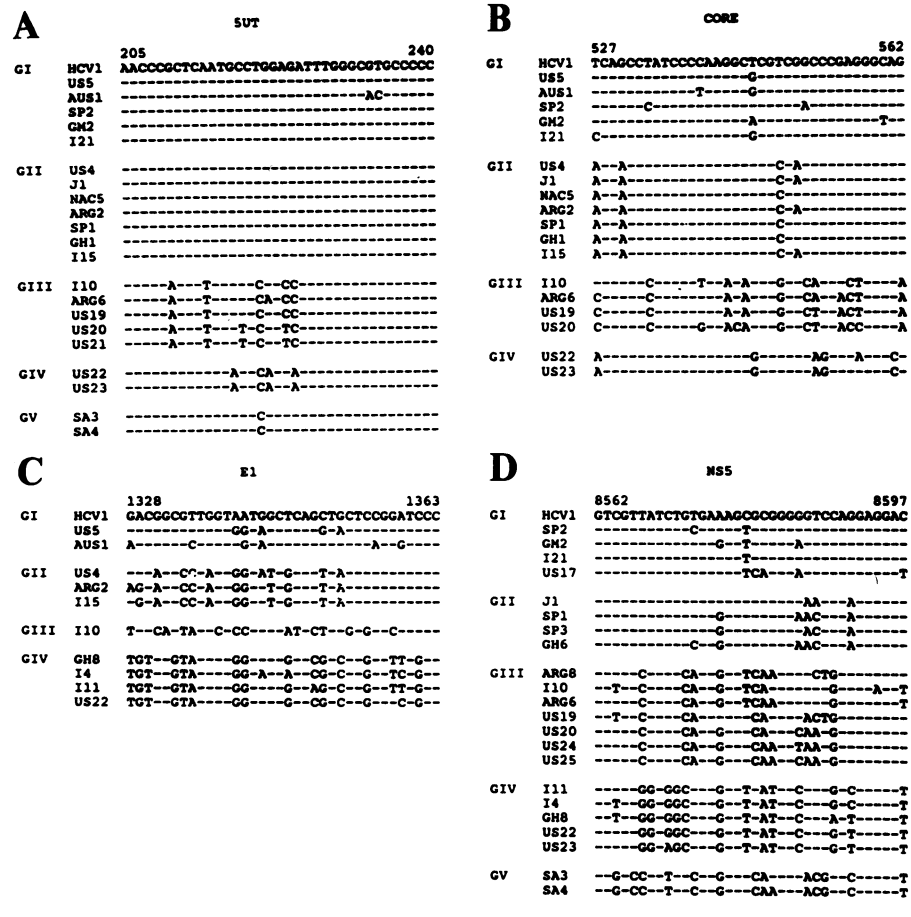

FIG. 2. Genotypic sequence motifs. The HCV1 sequence (4, 5) is shown at top, and the dashes indicate identical sequence compared with HCV1.

by using our definition (Fig. 2*D*). Among these 22 isolates, there are 4 from the United States, Spain, Germany, and Italy (isolates US17, SP2, GM2, and I21) assigned to GI, and 4 from Japan, Spain, and Germany (isolates J1, SP1, SP3, and GH6) assigned to GII. The remaining 14 isolates are distributed as follows: 7 isolates from the United States, Argentina, and Italy (isolates US19, US20, US24, US25, ARG6, ARG8, and I10) assigned to GIII, 5 isolates from the United States, Germany, and Italy (isolates US22, US23, GH8, I4, and I11) assigned to GIV, and 2 isolates from South Africa (SA3, SA4) assigned to GV. Note that the K1 and K2 types reported by Enomoto *et al.* (16) would be assigned to GII and GIII, respectively, based on their similar sequence motifs.

GI and GII Coexist Throughout the World. More extensive sequence was determined for 18 GI and GII HCV isolates from a wide geographic representation covering the 5′-end portion of the HCV genome (5UT, C, E1, and the N terminus of NS1/E2). Eight isolates—including HCV1, US5, US6, US17, AUS1, SP2, GM2, and I21—are GI, and 10 isolates—including US4, US18, J1, J14, NAC5, ARG2, SP1, GH1, GH6, and I15—are GII. Needleman and Wunsch (24) analysis was done on all sequences generated with GENALIGN. Not a single insertion or deletion was identified in any isolate examined when compared with HCV1. Although almost identical sequences were obtained for the 5UT region for all GI and GII isolates previously observed (5, 23), sequences downstream from 5UT clearly exhibited two distinct genotypes (Fig. 2). Despite observed genotypic variations, all isolates represent a single virus, HCV, due to their highly homologous 5UT sequences. One important observation was that both genotypes were found in geographically diverse regions, including Japan (12), Spain, Germany, Italy, and the United States.

In contrast to the 5UT region, which was highly homologous among all GI and GII isolates, a small hypervariable region located at the N terminus of NS1/E2 was found in both GI and GII. It should be noted that Weiner *et al.* (22) first reported a hypervariable domain among several GI HCV isolates, and Hijikata *et al.* (25) reported a similar hypervariable domain in several GII isolates.

Genotype-Specific Oligonucleotide Probes. Distinct patterns of nucleotide differences clearly delineated GI and GII. Based on these genotype-specific motifs, we designed four oligonucleotide probes in the C (GIC and GIIC) and E1 (GIE and GIIE) regions, which were specific for either genotype (see Table 1). A PCR fragment (40s/50a) was generated for all isolates that contained most of the C and E1 domains (see Fig. 1). Fig. 3*A* shows the results of agarose gel analysis, and Fig. 3 *B–E* shows the results of hybridization to probes GIC, GIIC, GIE, and GIIE, respectively. It can be seen that these probes possessed sufficient specificity to unambiguously distinguish these two genotypes. By designing specific nucleic acid probes based on the sequence motifs pertaining to each assigned genotype, this method can be generally applied as a rapid screening method for determining the prevalence of any genotype in a particular geographic area.

Amino Acid Sequence Variation. Amino acid sequence analysis revealed no interruption in the long open reading frame in any isolate sequenced. Less heterogeneity was observed in amino acid sequences than that in the nucleotide sequences due to a preponderance of nucleotide variations in the third-codon position. However, two distinct types can still be distinguished on the basis of amino acid sequence. Nucleotide and amino acid homologies among various domains of the two HCV genotypes were tabulated in Table 2. For amino acid sequence, the C domain was highly conserved among all GI and GII isolates (96–100%). The E1 domain was 90–98% homologous among isolates in the same genotype but

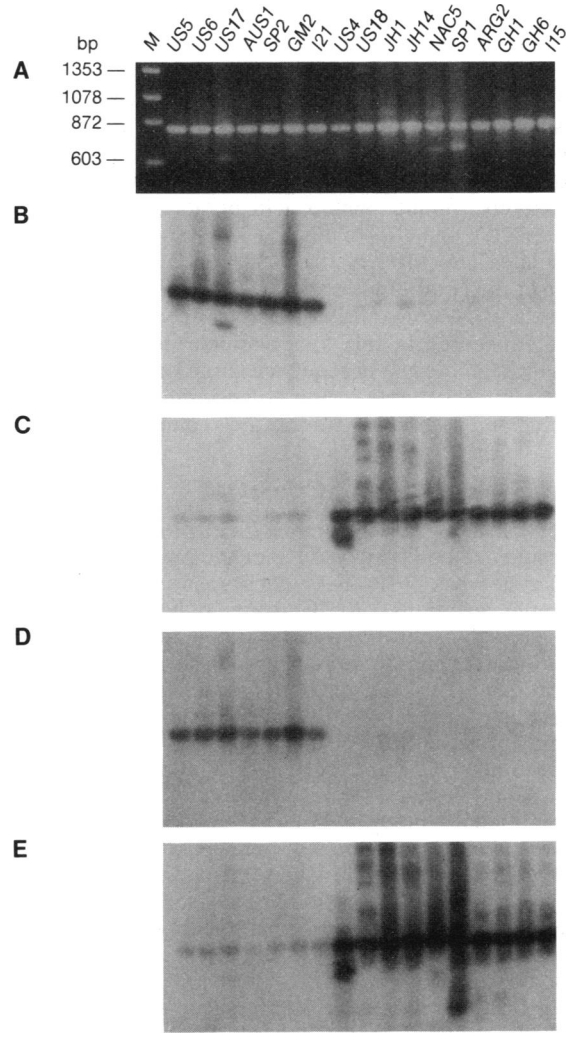

FIG. 3. Agarose gel and genotyping analyses. (*A*) Agarose gel analysis of the PCR fragment (40s/50a) generated for all 17 isolates. Four identical gels as in *A* are blotted onto nylon membranes and hybridized by ^{32}P-labeled oligonucleotide probes. The autoradiographs in *B–E* are with probes GIC, GIIC, GIE, and GIIE, respectively. Blots in *B* and *C* were washed with 75 mM sodium chloride/7.5 mM sodium citrate/0.1% SDS at 60°C; blots in *D* and *E* were washed with 150 mM sodium chloride/15 mM sodium citrate/0.1% SDS at 60°C.

Table 2. Nucleotide and amino acid sequence homologies in various domains

	Sequence homology, %				
	5UT	C	E1	NS1/E2	NS5
GI vs. GI	99–100	97–98 (98–99) [99]	91–96 (92–98) [98–100]	78–89 (63–82) [81–93]	95–98 (99–100) [99–100]
GI vs. GII	98–100	90–92 (96–97) [98–99]	73–76 (76–79) [93–96]	62–70 (56–69) [75–85]	79–82 (86–88) [92–94]
GII vs. GII	99–100	97–99 (98–100) [99–100]	92–94 (90–95) [97–99]	76–79 (68–75) [81–87]	89–93 (93–99) [95–99]

Amino acid-sequence homology is shown in parentheses. Numbers in brackets represent amino acid-sequence homology calculated by using a reduced set based on similarities in physical and chemical properties of the amino acids (26). Nucleotide positions for various domains were defined as 1–341 for 5UT, 342–914 for C, 915–1490 for E1, 1491–1694 for the N terminus of NS1/E2, and 8276–8615 for a part of NS5 (6).

only exhibited 76–79% homology among isolates between different genotypes. Table 2 shows, however, that the homology of the E1 domain between two genotypes increased to 93–96% when conserved replacement of amino acid residues was considered (26). The homology among isolates between GIII and GI was calculated at 91% and 71–75% from the partial sequences determined in the C and NS5 regions, respectively. Similar amino acid homology was also seen between GIV and GI isolates (71–73%) and between GV and GI isolates (75–76%) in the NS5 region. A hypervariable region at the N terminus of NS1/E2 was seen in both GI and GII isolates. Among all isolates examined, no identical sequence appeared in this hypervariable domain. To understand the significance of sequence variation in the E1 domain and of a hypervariable region downstream, detailed epitope mapping and serotyping studies are required.

DISCUSSION

Our conclusions can be summarized as follows: (*i*) Nucleotide sequence comparisons depict the existence of multiple HCV genotypes. (*ii*) These genotypes coexist in various geographic locations. (*iii*) The degree of heterogeneity varies with each domain. (*iv*) Our data confirm and extend a previous observation that a hypervariable region exists in the N terminus of NS1/E2 for both GI and GII.

We found several genotypes coexisting in various geographic locations—for instance, GI–GIV all found in the United States, GII and GIII in Argentina, GI and GII in Spain, GI, GII, and GIV in Germany, and GI–GIV in Italy. One additional genotype (GV) was found only in South Africa to date. Consistent with our findings, the coexistence of genotypes was reported in Japan (12, 14, 16) and France (27). To discern whether all genotypes coexist globally or whether there is a geographically dominant genotype would require screening more samples. In Japan, GII has been reported as the dominant genotype (14, 28).

Many factors contribute to the success or failure of the RT–PCR method. For instance, the RNA integrity depends on the efficiency of inhibiting RNase in the serum sample during the RNA isolation processes. In addition, the ability to produce DNA material suitable for sequencing depends not only on the efficiency of the RT reaction but also on the success of the PCR amplification. More importantly, the efficiency of PCR amplification depends ultimately on primer design. In this study, we have identified some seropositive samples that gave negative results with RT–PCR. We also identified some samples that were positive in the 5UT region by PCR but were not positive with any other primer pairs. Even though we identified four genotypes in the 5′-end region and five genotypes in the NS5 region, it cannot be precluded that additional genotypes exist and are yet to be discovered.

The assignment of HCV genotypes is useful but relatively arbitrary. When only the 5UT region is considered, fewer distinct genotypes can be assigned. On the other hand, when only the hypervariable region is considered, each isolate will represent a unique genotype. We should emphasize that long segments of sequence in several domains should be examined together for each isolate to classify its genotype. The functional significance of the sequence conservations or alterations within genotypes is unclear to date. An understanding of the sequence variation, however, is of paramount importance in the development of reliable diagnostic tests and vaccines and for epidemiological studies.

Note Added in Proof. We have collaborated with Dr. Gerald Myers at the Los Alamos National Laboratory for a phylogenetic analysis of the various HCV sequences. The preliminary results are consistent with our genotype assignment. The full account of this study will be reported elsewhere.

We thank our many friends and colleagues at Chiron and Daiichi Pure Chemicals for valuable discussions and Georgianne Trosin for aid in preparing the manuscript. We also thank Amy Weiner for critical review of the manuscript. This study was supported by Chiron Corporation and Daiichi Pure Chemicals.

1. Choo, Q.-L., Kuo, G., Weiner, A. J., Overby, L. R., Bradley, D. W. & Houghton, M. (1989) *Science* **244**, 359–362.
2. Kuo, G., Choo, Q.-L., Alter, H. J., Gitnick, G. I., Redeker, A. G., Purcell, R. H., Miyamura, T., Dienstag, J. L., Alter, M. J., Stevens, C. E., Tegtmeier, G. E., Bonino, F., Colombo, M., Lee, W.-S., Kuo, C., Berger, K., Shuster, J. R., Overby, L. R., Bradley, D. W. & Houghton, M. (1989) *Science* **244**, 362–364.
3. Choo, Q.-L., Weiner, A. J., Overby, L. R., Kuo, G., Houghton, M. & Bradley, D. W. (1990) *Br. Med. Bull.* **46**, 423–441.
4. Choo, Q.-L., Richman, K. H., Han, J. H., Berger, K., Lee, C., Dong, C., Gallegos, C., Coit, D., Medina-Selby, A., Barr, P. J., Weiner, A. J., Bradley, D. W., Kuo, G. & Houghton, M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 2451–2455.
5. Han, J. H., Shyamala, V., Richman, K. H., Brauer, M. J., Irvine, B., Urdea, M., Tekamp-Olson, P., Kuo, G., Choo, Q.-L. & Houghton, M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 1711–1715.
6. Houghton, M., Weiner, A., Han, J., Kuo, G. & Choo, Q.-L. (1991) *Hepatology* **14**, 381–388.
7. Kato, N., Hijikata, M., Ootsuyama, Y., Nakagawa, M., Ohkoshi, S., Sugimura, T. & Shimotohno, K. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 9524–9528.
8. Takamizawa, A., Mori, C., Fuke, I., Manabe, S., Murakami, S., Fujita, J., Onishi, E., Andoh, T., Yoshida, I. & Okayama, H. (1991) *J. Virol.* **65**, 1105–1113.
9. Okamoto, H., Okada, S., Sugiyama, Y., Kurai, K., Iizuka, H., Machida, A., Miyakawa, Y. & Mayumi, M. (1991) *J. Gen. Virol.* **72**, 2697–2704.
10. Takeuchi, K., Yoshihiro, K., Boonmar, S., Watanabe, Y., Katayama, T., Choo, Q.-L., Kuo, G., Houghton, M., Saito, I. & Miyamura, T. (1990) *J. Gen. Virol.* **71**, 3027–3033.
11. Takeuchi, K., Kubo, Y., Boonmar, S., Watanabe, Y., Katayama, T., Choo, Q.-L., Kuo, G., Houghton, M., Saito, I. & Miyamura, T. (1990) *Nucleic Acids Res.* **18**, 4626.
12. Okamoto, H., Okada, S., Sugiyama, Y., Yotsumoto, S., Tanaka, T., Yoshizawa, H., Tsuda, F., Miyakawa, Y. & Mayumi, M. (1990) *Jpn. J. Exp. Med.* **60**, 167–177.
13. Kubo, Y., Takeuchi, K., Boonmar, S., Katayama, T., Choo, Q.-L., Kuo, G., Weiner, A. J., Bradley, D. W., Houghton, M., Saito, I. & Miyamura, T. (1989) *Nucleic Acids Res.* **17**, 10367–10372.
14. Tsukiyama-Kohara, K., Kohara, M., Yamaguchi, K., Maki, N., Toyoshima, A., Miki, K., Tanaka, S., Hattori, N. & Nomoto, A. (1991) *Virus Genes* **5**, 243–254.
15. Takeuchi, K., Boonmar, S., Kubo, Y., Katayama, T., Harada, H., Ohbayashi, A., Choo, Q.-L., Kuo, G., Houghton, M., Saito, I. & Miyamura, T. (1990) *Gene* **91**, 287–291.
16. Enomoto, N., Takada, A., Nakao, T. & Date, T. (1990) *Biochem. Biophys. Res. Commun.* **170**, 1021–1025.
17. Chomczynski, P. & Sacchi, N. (1987) *Anal. Biochem.* **162**, 156–159.
18. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), 2nd Ed.
19. Steinhauer, D. A. & Holland, J. J. (1987) *Annu. Rev. Microbiol.* **41**, 409–433.
20. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988) *Science* **239**, 487–491.
21. Keohavang, P. & Thilly, W. G. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 9253–9257.
22. Weiner, A. J., Brauer, M. J., Rosenblatt, J., Richman, K. H., Tung, J., Crawford, K., Bonino, F., Saracco, G., Choo, Q.-L., Houghton, M. & Han, J. H. (1991) *Virology* **180**, 842–848.
23. Cha, T.-A., Kolberg, J., Irvine, B., Stempien, M., Beall, E., Yano, M., Choo, Q.-L., Houghton, M., Kuo, G., Han, J. H. & Urdea, M. C. (1991) *J. Clin. Microbiol.* **29**, 2528–2534.
24. Needleman, S. B. & Wunsch, C. D. (1970) *J. Mol. Biol.* **48**, 443–453.
25. Hijikata, M., Kato, N., Ootsuyama, Y., Nakagawa, M., Ohkoshi, S. & Shimotohno, K. (1991) *Biochem. Biophys. Res. Commun.* **175**, 220–228.
26. Miyata, T., Miyazawa, S. & Yasunaga, T. (1979) *J. Mol. Evol.* **12**, 219–236.
27. Li, J.-S., Tong, S.-P., Vitvitski, L., Lepot, D. & Trepo, C. (1991) *Gene* **105**, 167–172.
28. Nakao, T., Enomoto, N., Takada, N., Takada, A. & Date, T. (1991) *J. Gen. Virol.* **72**, 2105–2112.